

SIE 2016, 10º Simposio de Informática en el Estado

Interoperabilidad en Datos Biométricos: Arquitectura basada en estándares en una BDOR

Graciela Etchart, Carlos Alvez, Silvia Ruiz, Ernesto Miranda, Marcelo Benedetto, Juan José Aguirre, Santiago Delfin, Mauro Herlein,
Facultad de Ciencias de la Administración - Universidad Nacional de Entre Ríos
Av. Tavella 1424, Concordia, Entre Ríos - CP 3200
{getchart, caralv, silruiz, emiranda, marben,
juaagu}@fcad.uner.edu.ar {toti.delfin, herlein.mauro}@gmail.com

Resumen. Los organismos públicos y privados necesitan contar con un alto nivel de seguridad en distintas áreas. Las organizaciones estatales que emplean esta tecnología necesitan consultar los datos biométricos de manera eficiente e intercambiar los mismos con otros organismos asociados. De allí la necesidad de contar con modelos de datos y estándares que permitan la interoperabilidad entre sistemas y así facilitar la búsqueda e intercambio de datos. El objetivo de este proyecto, es desarrollar una arquitectura genérica utilizando la tecnología de las Bases de Datos Objeto-Relacionales (BDOR), acorde a normas internacionales, para la identificación de personas mediante el reconocimiento biométrico. Como prueba de esta arquitectura genérica se realizó la implementación del rasgo biométrico de iris en una BDOR. Esta implementación permite el reconocimiento biométrico mediante iris y la interoperabilidad entre organismos de manera eficiente y segura.

1 Introducción

Los organismos públicos y privados necesitan contar con un alto nivel de seguridad en distintas áreas para proteger tanto bienes materiales, como información de diverso tipo, ya sea en formato digital o de otra índole. Generalmente, el control de acceso a esta información, se efectúa a través de mecanismos tradicionales, como pueden ser, las claves de acceso y tarjetas magnéticas. Sin embargo, es posible que aparezcan fraudes aun con estos procedimientos: las tarjetas magnéticas se pueden perder o sustraer, las claves se pueden olvidar o adivinar, entre otros.

En este sentido, una solución alternativa o complementaria a las mencionadas es la utilización de sistemas biométricos. En los últimos años, las tecnologías biométricas han tenido un gran avance en su desarrollado. Esto se ha visto en los diferentes rasgos más utilizados (huella dactilar, iris, voz, rostro, etc.), así como también desarrollo de nuevos (por ejemplo, las venas de la mano) y la multibiometría [1].

Además de la importancia en la seguridad, la biometría juega un rol fundamental en la implementación de políticas públicas como las de gobierno electrónico y de políticas sociales [2]. Por esto, es importante que los organismos se apoyen en los sistemas

de reconocimiento biométricos para identificar a las personas, autenticar su identidad (pasaportes, documento de identidad), acceder a sistemas informáticos y reforzar la seguridad en accesos a áreas físicas.

Los organismos públicos que emplean esta tecnología necesitan consultar los datos biométricos de manera eficiente e intercambiar los mismos con otros organismos asociados. En este sentido, la adopción de estándares para el intercambio de datos biométricos posibilita la interconsulta con otros organismos (locales y extranjeros) para apoyar a las políticas de seguridad públicas, nacionales e internacionales. De allí la necesidad de contar con modelos de datos y estándares que permitan la interoperabilidad entre sistemas y así facilitar la búsqueda e intercambio de datos.

Por lo antes expuesto, en este trabajo se plantea como objetivo general desarrollar una infraestructura genérica en Bases de Datos Objeto-Relacionales [3] para la identificación de personas mediante el reconocimiento biométrico que permita la interoperabilidad entre organismos. Esta arquitectura podrá ser utilizada en diferentes organismos públicos que necesiten intercambios de información biométricas. Por ejemplo, seguridad social, servicios penitenciarios, servicios migratorios y otros servicios de seguridad, etc.

En las secciones siguientes se tratarán: en la sección 2 se describen los estándares adoptados, en la sección 3 se presenta la arquitectura propuesta y la implementación de un modelo en BDOR. Luego, en la sección 4 se presenta el modelo propuesto y por último, se presentan las conclusiones y trabajos futuros.

2 Estándares orientados a la interoperabilidad

Un importante avance en el logro de acuerdos de intercambio de datos, es el estándar ANSI/NIST ITL 1-2011, norma biométrica publicada en noviembre de 2011 y actualizada en 2013 [4]. Esta norma define cómo trabajar para garantizar la interoperabilidad de datos biométricos entre los distintos sistemas. Especifica el contenido, el formato y las unidades de medida para el intercambio de huellas dactilares, palmares, plantares, faciales, el iris, el ácido desoxirribonucleico (ADN), y otras muestras biométricas y datos forenses que pueden ser utilizados en el proceso de identificación o verificación de una persona. Es el estándar más utilizado por entes estatales, en gran medida debido a la intervención del Estado Nacional, que ha establecido el cumplimiento de los estándares internacionales y la adopción de prácticas adecuadas en la materia [5].

El estándar ANSI/NIST ITL 1-2011 define la composición de los registros de una transacción que pueden ser intercambiados. Los registros pueden ser de diferentes Tipos. Toda transacción se compone de un registro Tipo-1 y uno o más registros de los Tipos 2 al 99. En otras palabras, se requiere un registro Tipo-1 para describir de manera general la transacción, y al menos un registro de datos de otro Tipo. Todos los registros pertenecientes a una única transacción se transmitirán juntos.

Las muestras biométricas capturadas en cualquier organización, pueden ser intercambiadas con otros organismos asociados. Para esto el sistema de gestión de datos biométricos debe soportar transacciones de acuerdo con el estándar ANSI / NIST ITL 1-2011, lo que significa que el sistema debe ser capaz de generar y/o utilizar transacciones que sean morfológica, sintáctica y semánticamente adecuadas a los requerimientos de este estándar.

Si bien la arquitectura genérica que se propone en este trabajo, es adaptable a cualquier rasgo biométrico, aquí se realizó una implementación para el caso específico de iris. Para el intercambio de datos de imágenes del iris se utiliza el registro Tipo-17 del estándar ANSI/NIST-ITL 1-2011. Entre las exigencias de conformidad de este registro, tiene además los requisitos para la representación y compresión de imágenes de iris. La compresión se aplica a las imágenes para reducir su tamaño de almacenamiento y los tiempos de transmisión. Los algoritmos de compresión que el estándar acepta son: JPEG 2000 (Estándar ISO/IEC 15444:2004) [6] y PNG (Estándar ISO/IEC 15948:2004) [7].

2.1 JPEG 2000

El estándar internacional JPEG 2000 es un sistema de codificación de imágenes, que permite una compresión, transmisión y almacenamiento eficiente de imágenes fijas y de secuencias de imágenes. Este estándar describe un sistema de compresión de imágenes que tiene una gran flexibilidad, no sólo para la compresión de imágenes, sino también, para el acceso al *codestream*¹. Este último provee mecanismos para localizar y extraer datos de una porción de la imagen comprimida con el propósito de retransmisión, almacenamiento, visualización o edición. Este acceso permite almacenar y recuperar datos apropiados de la imagen comprimida para una aplicación dada. De esta manera, la división entre los datos de la imagen original y los de la imagen comprimida en varias formas permite extraer datos de imagen comprimida para formar una nueva imagen reconstruida con una resolución más baja o con menor precisión, o seleccionando regiones de interés de la imagen original. Esto hace posible adecuar un *codestream* a las características del canal de transmisión, del dispositivo de almacenamiento, o del dispositivo de visualización, independientemente del tamaño, el número de componentes, y la precisión de la imagen original.

2.2 PNG

El estándar internacional PNG, especifica un flujo de datos y un formato de archivo asociado para una imagen sin pérdida, comprimida, portátil y que puede ser transmitida a través de Internet. Soporta imágenes de color indexado, escala de grises y de color verdadero (RGB), con transparencia opcional, incluyendo además información del espacio de color e información auxiliar como textos de comentarios. Además, permite

¹ Secuencia de bits que representa la información de la imagen comprimida.

profundidades desde 1 a 16 bits, es robusto y provee chequeo de integridad de archivos y detección de errores comunes durante la transmisión. Entre las principales características del formato PNG se encuentra la portabilidad. Tanto la codificación, decodificación y transmisión deben ser independientes de las plataformas de software y hardware. El formato PNG fue ideado para su uso en Internet, por lo cual los procesos de codificación y decodificación pueden ser realizados en forma serial, a medida que el flujo de datos va siendo recibido sobre un canal de comunicación serial.

3 Arquitectura propuesta

La arquitectura propuesta para el intercambio de datos biométricos, aprovecha los beneficios establecidos en los requerimientos del estándar ANSI / NIST ITL 1-2011 en lo que respecta a la interoperabilidad de los sistemas.

Las tareas de generación, transmisión y recepción de registros de Tipo-17 del estándar, se encuentran contempladas dentro del sistema de intercambio de imagen del iris. Este sistema tiene un módulo que se encarga de la generación y transmisión de los registros necesarios para una transacción conforme al estándar y otro módulo para la recepción de estos registros a partir de organismos asociados (Fig. 1).

El registro de Tipo-17 de la citada norma sólo se utiliza para el intercambio de información. Sin embargo, también se debe considerar la generación y almacenamiento de metadatos necesarios para el reconocimiento, es decir, para los procesos de identificación y/o verificación de las personas a través del iris (IrisCode) [8][9], que incluye el subsistema de reconocimiento de iris [10].

Las imágenes que gestionan los sistemas mencionados se pueden obtener ya sea por operaciones con otros organismos o desde el sistema de captura. Para esto, se utiliza el sistema de captura y compresión de imagen de iris. Las estructuras de datos necesarias para implementar una base de datos biométrica son principalmente estructuras complejas. Esto puede causar algunos problemas cuando se trabaja con un modelo de base de datos relacional. Por este motivo, en este trabajo se optó por implementar la arquitectura en un modelo de Bases de Datos Objeto-Relacionales.

3.1 Modelo de datos Objeto-Relacional

La arquitectura propuesta incluye la representación de metadatos relacionados con los datos biométricos de iris, de modo que se puedan acceder, recuperar, comparar e intercambiar de una manera sencilla y eficiente. Además, como se mencionó anteriormente, se debe tener en cuenta la calidad de representación de las imágenes, en particular las imágenes comprimidas para el intercambio. Esto también se ajusta a las especificaciones de las normas adoptadas.

La complejidad de las estructuras de datos necesarias para implementar una base de datos biométrica puede causar algunos inconvenientes cuando se trabaja con un mode-

lo de base de datos relacional, a causa de algunas limitaciones de este modelo respecto a los siguientes aspectos:

1. Las estructuras de los metadatos necesarios, tanto para registros ANSI / NIST ITL 1-2011 y para la comparación de los datos biométricos (*IrisCode*), son estructuras no atómicas, por lo que en las bases de datos relacionales, con el fin de respetar la primera forma normal, se deben separar en varias tablas. Esto hace que su tratamiento sea menos eficiente. Por ejemplo, al comparar dos conjuntos de metadatos será necesario llevar a cabo operaciones de reunión (join).

2. El procesamiento de los datos debe realizarse desde lenguajes de propósito general (C ++, Java, Python, entre otros.), lo que presenta el problema de la compatibilidad y la dependencia de los lenguajes de programación. Esto se debe a que deben transferir los datos generados por los métodos implementados en lenguajes diferentes al de la base de datos. Este problema no es trivial, y debe ser tratado para cada lenguaje en particular.

3. Relacionado con lo anterior, la forma que un lenguaje de programación accede a los datos, puede traer problemas de eficiencia en el uso de redes. Esto puede verse en el caso de la adquisición de datos necesarios en los métodos de comparación iris (verificación/identificación). Esto es porque deben crearse cursores para transferir los Iris-Code a través de la red generando mucho tráfico innecesario, ya que este método devuelve un único valor lógico. Por ejemplo, en una comparación utilizando el proceso de verificación, retorna si verifica o no.

Teniendo en cuenta los inconvenientes mencionados de las bases de datos relacionales, aquí hemos optado por utilizarla tecnología Objeto-relacional. Esta tecnología puede proporcionar soluciones a las limitaciones de las bases de datos relacionales, ya que proporcionan las siguientes capacidades [11]:

- *Definir tipos de datos*: estos tipos pueden contener estructuras complejas como colecciones, objetos grandes, etc. sin la limitación de la primera forma normal.
- *Definir e implementar el comportamiento de los datos*: se pueden crear los métodos que gestionen los datos de los tipos definidos, y así facilitar el acceso seguro desde las aplicaciones que los utilicen.
- *Definir e implementar métodos de acceso de dominio*: esto permite mejorar los tiempos de acceso en la recuperación de datos de un dominio específico, como en el caso de los datos biométricos.

Las facilidades mencionadas, posibilitan crear infraestructuras que extienden los servicios del Sistema de Gestión de Base de Datos [11]. Estas le permiten a este sistema tratar con datos de un dominio específico, como pueden ser los datos biométricos; críticos en los programas de identidad y seguridad.

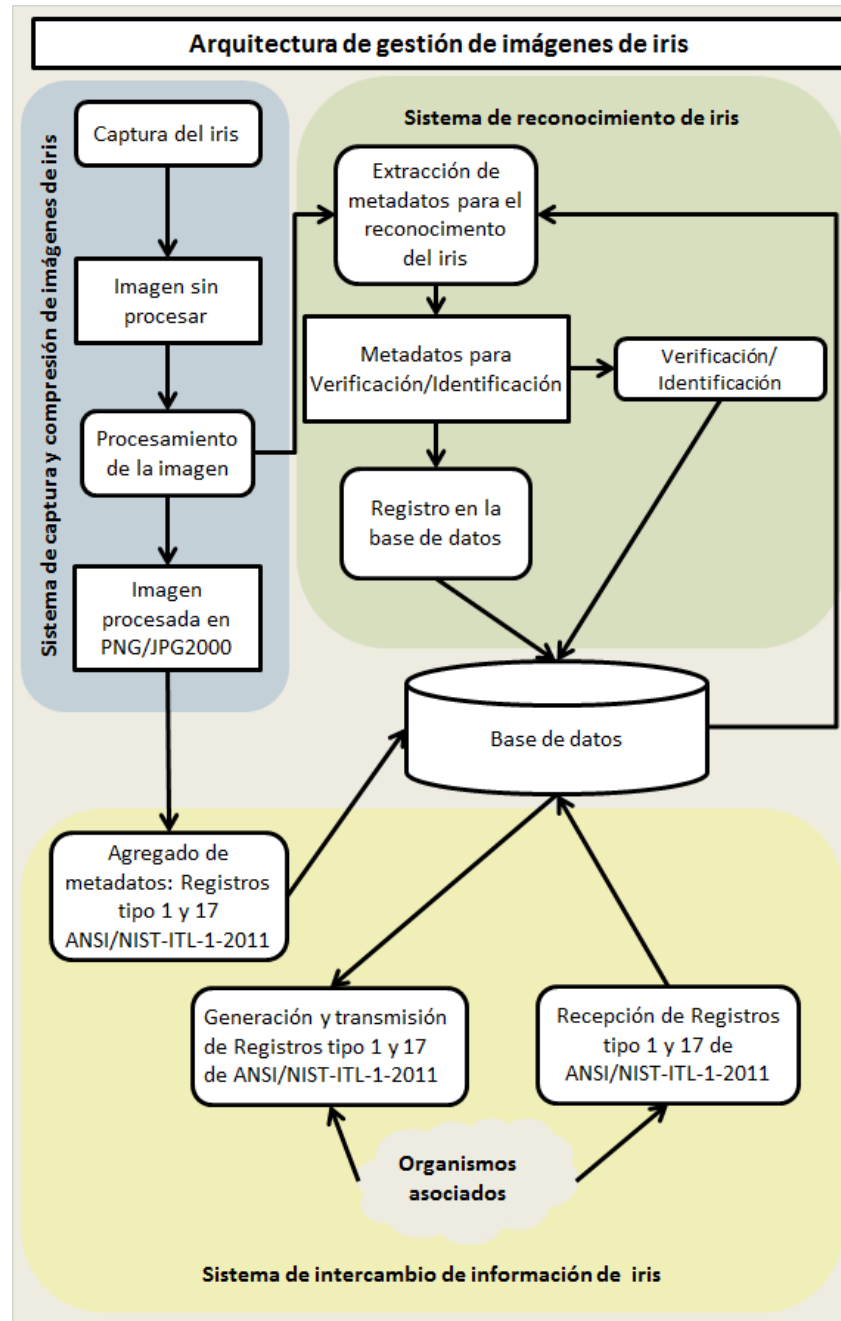


Fig. 1: Arquitectura de gestión de imágenes de iris.

En Alvez, et. al.[10] se presenta un Diagrama de clase UML para los metadatos requeridos por el registro Tipo-17 del estándar ANSI/NIST-ITL 1-2011. En ese diagrama se representan las clases con los atributos necesarios para la generación del registro de transacción. Estos atributos, en general, se tratan de estructuras compuestas o repetitivas que representan diferentes tipos de información sobre las imágenes capturadas, dispositivos empleados en la captura y datos referentes a sitios y fechas de captura.

Por otro lado, en el diagrama de clases UML de la Fig. 2 se observan las clases de captura y almacenamiento de las imágenes con sus respectivos atributos. Las clases IMGJPG2000 e IMGPNG heredan de la clase ImagenProcesada y tienen una relación de asociación con los paquetes OpenJPEG² y LIBPNG³.

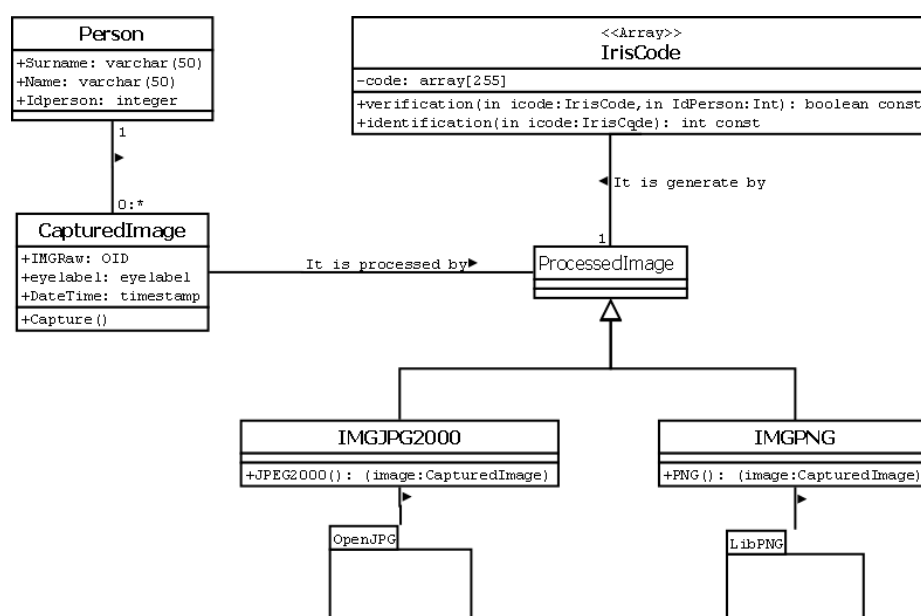


Fig. 2. Diagrama de clases con los tipos de datos necesarios para la captura y procesamiento de la imagen y el sistema de reconocimiento del iris.

La clase ImagenProcesada genera dos subclases por herencia de acuerdo a su implementación, las cuales se asocian con sus respectivos paquetes de librerías. Esta clase, a su vez, genera una clase IrisCode con los respectivos métodos de verificación e identificación.

En la siguiente sub-sección, se presentarán los detalles de la implementación del modelo en una Base de Datos Objeto-Relacional.

² <http://www.openjpeg.org/>

³ <http://www.libpng.org/pub/png/libpng.html>

3.2 Implementación en una BDOR

Dado el modelo de datos planteado para la arquitectura propuesta, la implementación en una BDOR debe incluir:

- La creación de tipos específicos para datos biométricos: Esto incluye los tipos de datos para datos crudos (imágenes del iris en nuestro caso), tipos de para el código del iris (*IrisCode*), y todos los metadatos necesarios para generar registros de transacción de la norma ANSI / NIST ITL1-2011.
- Definir e implementar los métodos que gestionan los tipos previamente definidos: Esto incluye, entre otros, la generación de codificaciones específicas, métodos de comparación (verificación/identificación) y los métodos de generación e importación de registros ANSI / NIST ITL1-2011.
- La creación de índices de dominio para mejorar el tiempo de respuesta en los procesos de identificación: en la identificación biométrica, es necesario acceder a múltiples registros para reconocer a una persona. La mejora de la eficiencia en este acceso no es fácil de lograr con los índices tradicionales, como los árboles B, hash, etc. Esto es debido a que los códigos generados (plantillas) para la comparación, son datos multidimensionales, por lo que hace necesario métodos de acceso específicos para este dominio.

Para esta implementación de la base de datos se analizaron motores de Bases de Datos Objeto-Relacionales tomando como referencia dos de las alternativas más populares, Oracle y PostgreSQL. Si bien Oracle cuenta con un mayor grado de madurez, se optó por PostgreSQL por ser una herramienta de software libre y dar soporte a todos los requerimientos del proyecto.

Se procedió a la descarga e instalación del motor de base de datos PostgreSQL en su versión 9.4 (última versión estable al momento) sobre ambientes Windows y Linux. Previo a esto se instalaron los sistemas operativos Windows 7 de 64 bits y Linux Mint 17 de 64 bits y de 32 bits en las computadoras del gabinete de acuerdo a las características de hardware de las mismas. También se realizó la instalación de las herramientas básicas para las necesidades que pudieran surgir durante la tarea de codificación.

Dada la necesidad de que cualquier modelo que se adopte para la base de datos propia pueda generar un registro de Tipo-17 del estándar ANSI-NIST ITL 1 - 2011 se decidió por comenzar a codificar la estructura de este modelo.

Para la implementación del modelo se tomó el diagrama de clases UML como base para la codificación de la estructura del registro de Tipo-17 en el lenguaje de procedimientos propio de PostgreSQL llamado PL/pgSQL.

Primero se centró el trabajo en la codificación de las estructuras más básicas y en sucesivas iteraciones se fueron refinando los controles referidos a restricciones en los valores de los campos.

Para desarrollar las funciones definidas se analizaron varios lenguajes de programación que son soportados por PostgreSQL. Dicho motor no tiene la forma de inter-

prestar el código de una función en un lenguaje procedural, entonces un manejador se encarga de delegar la tarea al intérprete del lenguaje. El soporte de lenguajes de este motor permite utilizar C, Tcl, Perl y Python, entre otros lenguajes, para la creación de estas funciones. Finalmente, dadas las características de código abierto, la eficiencia en las estructuras de datos, el enfoque a la programación orientada a objetos, se optó por desarrollar las funciones correspondientes en el lenguaje Python y las extensiones de este lenguaje con el motor PostgreSQL.

Con esta implementación se logró extender los servicios prestados por el sistema de gestión de base de datos PostgreSQL para gestionar datos biométricos basado en el estándar ANSI / NIST ITL1-2011, para el caso específico de iris. La implementación del modelo presentado en la Fig. 2 contempla los metadatos y métodos necesarios para el reconocimiento por el rasgo de iris.

Conclusiones y trabajos futuros

Las organizaciones, tanto públicas como privadas, se tornarán cada vez más dependientes de herramientas de manejo de la identidad más robustos. La biometría jugará un rol clave en atender los nuevos desafíos.

La solución propuesta en este proyecto apunta a obtener una mayor fiabilidad, rendimiento y universalidad, utilizando un sistema biométrico de reconocimiento de iris, basado en estándares. Es decir, un desarrollo personalizado y ajustado a los estándares que permita captura, procesar, almacenar e intercambiar datos y garantice la interoperabilidad entre distintas áreas y/o aplicaciones.

En etapas posteriores se trabajará en la incorporación de otro rasgo biométrico y se estudiarán alternativas para la creación de índices de dominio que permitan mejorar los tiempos de respuestas en los procesos de reconocimiento.

Referencias

1. M. Ghayoumi, "A review of multimodal biometric systems: Fusion methods and their applications," Computer and Information Science (ICIS), 2015 IEEE/ACIS 14th International Conference on, Las Vegas, NV, 2015, pp. 131-136
2. Delfin, Santiago; Miranda, Ernesto; Ruiz, Silvia. Orientadores: Alvez Carlos Eduardo; Etchart Graciela. "Biometría como herramienta clave de inclusión social". AUGM, XXIII Jornada de Jóvenes Investigadores, del 25 al 27 de agosto del 2015, La Plata.
3. Melton Jim, "(ISO-ANSI Working Draft) Foundation (SQL/Foundation)", ISO/IEC 9075-2:2003 (E), United States of America (ANSI), 2003.
4. Wing B. ANSI/NIST-ITL 1-2011. Update: 2013. Information Technology: American National Standard for Information Systems Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information. December 2013.
5. Julio Fuoco, "Tendencias Biométricas, desafíos y oportunidades". En Biometrías 2. Jefatura de Gabinete de Ministros. Presidencia de la Nación, 2011.

6. ISO/IEC 15444-1:2004 Information Technology-JPEG 2000 Image Coding System: Core Coding System, Second Edition, 2004-09-15.
7. ISO/IEC 15948:2004 Information Technology-Computer Graphics and Image Processing-Portable Network Graphics (PNG): Functional Specification, First Edition, 2004-03-01.
8. Daugman, J.: "High condence visual recognition of persons by a test of statistical independence". IEEE Transaction Pattern Analysis and Machine Intelligence 15(11), 1148-1161 (Novem-ber 1993).
9. J. Daugman and C. Downing, "Effect of severe image compression on iris recognition performance," IEEE Transactions on Information Forensics and Security, vol. 3, no. 1, p. 52-61, 2008.
10. Carlos Alvez, Graciela Etchart, Silvia Ruiz, Ernesto Miranda, Juan Aguirre, Marcelo Benedetto, Santiago Delfin. "Iris Information Management in Object-Relational Databases". XXI Congreso Argentino en Ciencias de la Computación (CACIC 2015), Junin
11. Carlos E. Alvez, Aldo R. Vecchietti. "Combining Semantic and Content Based Image Retrieval in ORDBMS". KES'2010, Lecture Notes in Computer Science, Editors Rossitza Setchi. Ivan Jordanov, Robert J. Howlett, Lakhmi C. Jain. Springer-Verlag Berlin Heidelberg (2010), Volume 6277/2010, 44-53